Matthew effects in reading: a comparison of latent growth curve models and simplex models with structured means

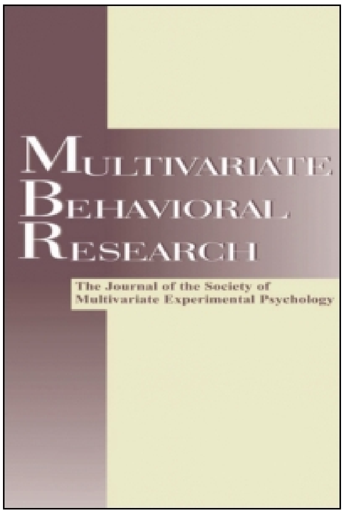
Bast, J.W.; Reitsma, P.

published in
Multivariate Behavioral Research
1997

DOI (link to publisher)
10.1207/s15327906mbr3202_3

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Bast, J. W., & Reitsma, P. (1997). Matthew effects in reading: a comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research*, *32*, 135-167. https://doi.org/10.1207/s15327906mbr3202_3



Multivariate Behavioral Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653673

Mathew Effects in Reading: A Comparison of Latent Growth Curve Models and Simplex Models with Structured Means

Janwillem Bast; Pieter Reitsma

Online publication date: 10 June 2010

To cite this Article Bast, Janwillem and Reitsma, Pieter(1997) 'Mathew Effects in Reading: A Comparison of Latent Growth Curve Models and Simplex Models with Structured Means', Multivariate Behavioral Research, 32: 2, 135 — 167
To link to this Article: DOI: 10.1207/s15327906mbr3202_3
URL: http://dx.doi.org/10.1207/s15327906mbr3202_3

Multivariate Behavioral Research, *32* (2), 135-167

Matthew Effects in Reading: A Comparison of Latent Growth Curve Models and Simplex Models with Structured Means

Janwillem Bast and Pieter Reitsma
Paedologisch Instituut — Vrije Universiteit Amsterdam

The Matthew effect hypothesis in reading predicts that the gap between good and poor readers increases with time. Although, intuitively appealing, the Matthew effect has hardly been empirically studied in longitudinal studies of reading. Two competing longitudinal models were used to represent the Matthew effect hypothesis: the Latent Growth Curve model and the Simplex model with structured means. It is argued that on the basis of theoretical and empirical arguments the Simplex model should be preferred to represent and analyze the Matthew effect hypothesis. However, the results of the Simplex models imply that conceptual refinement and clarification of Matthew effects in reading are needed.

Despite the effort and the many resources available in the schools of literate societies, it is clear that not all children become proficient readers. The principal difference between reading and other uses of language is that readers exercise their abilities in response to graphic rather than acoustic signals (Perfetti, 1985). Therefore, the first task of every novice reader is to decode letter strings into familiar sounds. Initially, the reading process can be defined as decoding or word recognition, that is aimed at the identification of a written word in order to access its meaning. However, the ultimate goal of every reader is not to identify the meaning of words in isolation, but to comprehend the meaning of text. Comprehension in reading can be referred to as the second reading process.

Evidently, there are large differences in the ability to profit from initial reading instruction and to acquire fluent reading skills. Furthermore, individual differences in reading ability seem to fan out with time. For example, it has been documented that differences among the reading achievement levels of twelfth graders are larger than differences among first

---

This research was supported by Grant SGW 575-91-002 from the Dutch Organization of Scientific Research (NWO) to the second author. The helpful comments of Connor Dolan, and two anonymous reviewers in preparing this article are gratefully acknowledged.

Address correspondence to Pieter Reitsma, Paedologisch Instituut - VU Amsterdam, P.O. Box 303, 1115 ZG Duivendrecht, The Netherlands (e-mail: P.Reitsma@psy.vu.nl).

graders (Daneman, 1991). Also, Williamson, Appelbaum and Epanchin (1991) found a steady increase in mean achievement scores for reading ability from Grade 1 to Grade 8 along with a constant increase in the variability of the achievement scores. An important issue then is how individual differences in reading ability develop with time.

Using the biblical analogy of the rich getting richer and the poor getting poorer the phenomenon of increasing achievement differences is occasionally referred to as Matthew effects. Following Merton (1968), who introduced the notion of Matthew effects to describe differences in scientific productivity, and Walberg and Tsai (1983), who found evidence for the effect in an educational setting, Stanovich (1986) used the concept of Matthew effects to describe and explain individual differences in the development of reading. The Matthew effect model of Stanovich can be described as a set of interrelated hypotheses. The first assumption concerns increasing performance differences between subjects in the course of reading development. This means that the development of individual variation in reading can be characterized by a stable rank ordering of individuals and an increase of performance differences among persons instead of stability or a decrease of these differences. The second assumption is that the phenomenon of increasing achievement differences is caused by (reciprocal) causal relations between reading and other cognitive skills, attitudes or behaviors. With respect to reading, reciprocal causation means "... that individual differences in a particular process may cause differential reading efficiency, but that reading itself may in turn cause further individual differences in the process in question" (Stanovich, 1986 p. 378). However, factors that enter into (reciprocal) causal relationships with reading ability can change with development. Not every relationship is assumed to be operative in similar fashion during each stage in the development of reading. Some relationships will be developmentally limited, that is that individual differences in a particular cognitive process may be a causal determinant of variation in reading achievement early in development, but at some point have no further effects on the level of reading efficiency. For instance, differences in word recognition skills lead to differences in reading comprehension. Students who are not preoccupied with decoding issues can give their full attention to the process of constructing meaning (Perfetti, 1985). Word-recognition skills may account for relatively little of the variance once readers get beyond the beginning stages of reading. After that, general language comprehension skills seem to be a larger source of individual differences in reading comprehension (Hoover & Gough, 1990).

The Matthew effect hypothesis, as put forward by Stanovich (1986), is primarily a model for individual differences in development. No explicit claims about the developmental mean curve are made. When the longitudinal mean structure and the covariance structure are modeled simultaneously, the assumption is often made that changes in the mean trend and changes in individual differences can be attributed to a common cause (Mandys, Dolan & Molenaar, 1994). However, this assumption may not be valid. Factors that influence the development of reading in general may or may not influence individual differences in reading ability. For example, circumstances that permit almost all children to run may not play any role in how well or fast one individual can run relative to another (McCall, 1981). In other words, the implicit assumption of common causation should be examined empirically (Mandys et al., 1994).

Although intuitively appealing, the Matthew effect has rarely been empirically documented in longitudinal studies of reading (Shaywitz et al., 1995). A possible reason for this is that the Matthew effect, as described by Stanovich (1986), is presented as a metaphoric model. The model consists mainly of verbal descriptions and there are no clear guidelines for its empirical investigation. In other words, the model lacks testability. In order to test this model or aspects of this model empirically, a formal redescription in terms of a mathematical or statistical model of the underlying growth processes is needed. In the present study two statistical growth models with different representations of Matthew effects in reading are compared: the Latent Growth Curve model (LGC) and the Simplex model with structured means.

Two Longitudinal Structural Equation Models

In any longitudinal study, two types of time dependent relations can be analyzed: patterns of means across individuals over time and patterns of variability across individuals over time. The interpretation of both type of patterns can be troublesome. For instance, when a pattern of increasing variation and increasing means across time is observed, differential trajectories of learning could account for this pattern. However, differential reliability of the observed variables could also act to produce a false pattern of increasing spread (Rudinger & Wood, 1989). Therefore, a clear distinction between structural and measurement models must be made. The general structural equation model makes such a distinction possible. This model consists of two parts: the measurement model and the structural equation model. The measurement model specifies the relation between observed variables and hypothetical constructs or latent variables. The structural equation model specifies the relationships among the latent variables, describes the effects, and assigns the explained and unexplained variance. A description of the general form and methods of structural equation models can be found in, for instance, Bollen (1989) or in the LISREL manual (Jöreskog & Sörbom, 1989).

The measurement model for y (discarding the subject index) of the general structural equation model with mean structures is defined by:

$$\mathbf{Y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{Y}$ is a $p \times 1$ vector of observed variables, in the present case the repeated measures, $\boldsymbol{\tau}$ is a $p \times 1$ vector of constant intercept terms, $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings, $\boldsymbol{\eta}$ is a $q \times 1$ vector of latent variables, and $\boldsymbol{\epsilon}$ is a $p \times 1$ vector of measurement errors in y. The structural part of the model is defined by:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\zeta} \tag{2}$$

where $\boldsymbol{\alpha}$ is a $q \times 1$ vector of constant intercept terms, and $\mathbf{B}$ is a $q \times q$ matrix of coefficients expressing the structural relationships between the $\boldsymbol{\eta}$ variables. Finally, $\boldsymbol{\zeta}$ is a $q \times 1$ vector of equation residuals or random disturbances. Under the assumptions that the $\boldsymbol{\epsilon}$s are uncorrelated among themselves, uncorrelated with $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$, $\boldsymbol{\eta}$ is uncorrelated with $\boldsymbol{\zeta}$, and $E(\boldsymbol{\epsilon}) = 0$ and $E(\boldsymbol{\zeta}) = 0$, these equations lead to the following mean structures and implied model covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{3}$$

$$\boldsymbol{\mu}_y = \boldsymbol{\tau} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} \tag{4}$$

$$\boldsymbol{\mu}_{\eta} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} \tag{5}$$

where $\boldsymbol{\Psi}$ is the covariance matrix ($q \times q$) of the equation residuals $\boldsymbol{\zeta}$, and $\boldsymbol{\Theta}$ is the covariance matrix ($p \times p$) of the measurement errors.

Within a structural equation modeling framework there are two types of longitudinal models that seem particularly suitable to represent Matthew effects in reading: models based on the common factor model and models based on a simplex structure. The common factor model enables individual growth modeling (Rogosa, Brandt & Zimowski, 1982; Rogosa & Willett, 1985b) by means of latent growth curve analysis (McArdle & Epstein, 1987; Meredith & Tisak, 1990; Willett & Sayer, 1994). Two latent factors, level and shape, represent dimensions of individual differences in growth over time. These common factors are used to describe the values of the status and growth of individuals on a trait at all occasions. Apart from individual differences in development, average growth of a group of subjects can be modelled simultaneously.

To specify the LGC model as a structural equation model with mean structures (Willett & Sayer, 1994), we set the **B** matrix and the $\boldsymbol{\tau}$ vector to zero. The model for individual growth curves is now defined by the measurement model for y:

$$\mathbf{Y} = \boldsymbol{\Lambda\eta} + \boldsymbol{\epsilon} \tag{6}$$

where the $\boldsymbol{\Lambda}$-matrix contains the empirical growth records, that is constants and the known times of measurement, the $\boldsymbol{\eta}$-matrix contains the estimated individual latent growth parameters, and $\boldsymbol{\epsilon}$ represents measurement error. The model for interindividual differences in change can be written as a structural model:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\zeta} \tag{7}$$

where the $\boldsymbol{\alpha}$-vector contains the population averages of the latent individual growth parameters, and as follows from Equation 7, $\boldsymbol{\zeta} = \boldsymbol{\alpha} - \boldsymbol{\eta}$, that is, $\boldsymbol{\zeta}$ is a latent residual vector that contains the deviations of the individual growth parameters from their respective population means. These equations lead to the following implied covariance matrix and mean structures:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Psi\Lambda}' + \boldsymbol{\Theta} \tag{8}$$

$$\boldsymbol{\mu}_y = \boldsymbol{\Lambda\alpha} \tag{9}$$

$$\boldsymbol{\mu}_{\boldsymbol{\eta}} = \boldsymbol{\alpha} \tag{10}$$

where the $\boldsymbol{\psi}$ covariance matrix contains the variances and covariances of the latent individual growth parameters.

The classical assumptions of the distribution of the measurement errors are homoscedasticity and independence. These assumptions may be too stringent in the case of longitudinal data. The reliability of the observed measures may not be constant over time. Moreover, when measurements are closely spaced in time, autocorrelations can arise. Therefore, for repeated measures data other specifications of the $\boldsymbol{\Theta}$-matrix, for instance, pairwise correlated or heteroscedastic errors, must be specified and tested.

If we want to specify, for instance, a latent quadratic growth curve model for six observed repeated measures, the individual growth curve model of the p^{th} person equals:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_{1p} \\ \mathbf{Y}_{2p} \\ \mathbf{Y}_{3p} \\ \mathbf{Y}_{4p} \\ \mathbf{Y}_{5p} \\ \mathbf{Y}_{6p} \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \\ 1 & t_4 & t_4^2 \\ 1 & t_5 & t_5^2 \\ 1 & t_6 & t_6^2 \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \pi_{0p} \\ \pi_{1p} \\ \pi_{2p} \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{1p} \\ \epsilon_{2p} \\ \epsilon_{3p} \\ \epsilon_{4p} \\ \epsilon_{5p} \\ \epsilon_{6p} \end{bmatrix}$$

where $t_1 \ldots t_6$ represent the known times of measurement, π_0, π_1, and π_2 represent the estimated latent growth parameters of respectively the intercept, the linear component, and quadratic component, and $\boldsymbol{\epsilon}$ represents measurement error. The model describing individual differences in growth now equals:

$$\boldsymbol{\zeta} = \begin{bmatrix} \pi_{0p} - \mu_{p0} \\ \pi_{1p} - \mu_{p1} \\ \pi_{2p} - \mu_{p2} \end{bmatrix}, \text{ and } \mathbf{\Psi} = \mathrm{Cov}(\boldsymbol{\zeta}) = \begin{bmatrix} \sigma_{p0}^2 & \sigma_{\pi0\pi1} & \sigma_{\pi0\pi2} \\ \sigma_{\pi1\pi0} & \sigma_{p1}^2 & \sigma_{\pi1\pi2} \\ \sigma_{\pi2\pi0} & \sigma_{\pi2\pi1} & \sigma_{\pi2}^2 \end{bmatrix}$$

where μ_0, μ_1, and μ_2 represent the population means of the latent individual growth parameters, and $\boldsymbol{\psi}$ is the covariance matrix of the latent individual growth parameters.

The simplex model is particularly well suited to longitudinal series in which there is occasion-to-occasion transmission, that is to say, that the observation at time 2 depends on the observation at time 1, and in turn, the observation at time 3 depends on the observation at time 2, and so forth. This pattern is well described by a first-order autoregressive or Markov process. When measurement error of the observed variables is included in the model the simplex property shifts to the latent level. This model is known as the quasi simplex. Jöreskog (1970, 1979; Jöreskog & Sörbom, 1989; Werts, Linn & Jöreskog, 1978) formulated the quasi simplex model in the framework of structural equation models. Simplex models are primarily designed to analyse individual differences in development but they can be extended to include the simultaneous analysis of the means. The model thereby addresses both the trend of the average growth curve as well as the stability of individual differences.

In the quasi simplex model (QSM) the observed variables at time $i(\mathbf{Y}_i)$ are assumed to be related to their corresponding latent variables ($\boldsymbol{\eta}_i$) and measurement error ($\boldsymbol{\epsilon}_i$) by the equation:

(11) $$\mathbf{Y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$$

with $\boldsymbol{\Lambda} = \mathbf{I}$ for the one-indicator quasi simplex model. The simplex structure of the latent variables can be stated as:

(12) $$\boldsymbol{\eta}_{i+1} = \beta_{i+1}\boldsymbol{\eta}_i + \zeta_{i+1}$$

where $\boldsymbol{\beta}_{i+1}$ is the regression coefficient in the regression of $\boldsymbol{\eta}_{i+1}$ on $\boldsymbol{\eta}_i$, ζ_{i+1} is the equation residual, and $\boldsymbol{\eta}_1 = \zeta_1$ at the start of the time series.

The quasi simplex model is extended to include structured means in the following manner (Hanna & Lei, 1985; Mandys et al., 1994). The observed mean at time i is the sum of a constant $\boldsymbol{\tau}$, an arbitrary intercept term, and the mean of the latent variable $\boldsymbol{\eta}_i$:

(13) $$\mu_{yi} = \boldsymbol{\tau} + \mu_{\eta i}$$

(14) $$\mu_{yi+1} = \boldsymbol{\alpha}_{i+1} + \beta_{i+1}\mu_{\eta i}$$

These equations lead to the following mean structures and implied model covariance matrix $\boldsymbol{\Sigma}$:

(15) $$\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B}')^{-1} + \boldsymbol{\Theta}$$

(16) $$\mu_y = \boldsymbol{\tau} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha}$$

(17) $$\mu_\eta = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha}$$

The one-indicator quasi simplex model is not identified without additional constraints (Jöreskog, 1970). In particular, the parameters of the first and last occasion cannot be uniquely estimated. The most straightforward way to solve this indeterminacy is to assume that the error variances $\boldsymbol{\epsilon}$ of the first two occasions are equal and that the error variances of the last two occasion are equal (Jöreskog & Sörbom, 1989). However, other error structures are possible and testable, as long as a suitable reduction of the number of parameters is obtained.

Moreover, also the model for the means is not identified because there are more parameters than observed means. Thus, a reduction of the number of parameters is required. One possibility to model the mean trend is to take the regression of the latent mean $E[\boldsymbol{\eta}_{i+1}]$ on $E[\boldsymbol{\eta}_i]$, and to specify $\boldsymbol{\alpha}_{i+1}$ as constant throughout the time series (Mandys et al., 1994). We then arrive at:

(18) $$\mu_{\eta 1} = \alpha$$

(19) $$\mu_{\eta i+1} = \alpha + \beta_{i+1,i}\eta_i$$

Within a structural equation modeling framework three different but closely related hypotheses of Matthew effects in reading must be represented. First, statistical models that are candidates for models of growth in terms of Matthew effects in reading should incorporate the expected finding of increasing interindividual variance in combination with a high stability of the rank ordering of individuals. Second, the selected statistical growth models should also incorporate the expectation that different factors affect the interindividual variance in reading ability at different times, that is, influences change with development. Third, because the Matthew effect hypothesis of Stanovich (1986) is presented as a model for individual differences, the assumption that both individual differences and the developmental mean trend are not the result of an unspecified common cause should be tested.

The latent growth curve model and the quasi simplex model represent different conceptualizations of growth in reading. An example of a LGC and a Quasi simplex model for four waves of data is depicted in Figure 1.

As can be derived from Figure 1, in the simplex model differences in growth are not only a function of prior standing (the ηs) but also of random disturbances (the ζs). Because these random disturbances are uncorrelated with previous status on the underlying construct, initial status and variables influencing initial status have a decreasing influence. This means that other source of variance than those responsible for initial performance differences can be incorporated into the model. This feature makes the simplex model well suited to represent the expectation that different factors affect the interindividual variance at different times. However, as a consequence of this random variance the rank orderings will change. This can easily be demonstrated by looking at the covariance matrix implied by the quasi simplex model:

(20) $$\mathrm{var}(y_i) = \mathrm{var}(\eta_i) + \mathrm{var}(\epsilon_i)$$

(21) $$\mathrm{var}(\eta_{i+1}) = \beta^2_{i+1,i}\mathrm{var}(\eta_i) + \mathrm{var}(\zeta_{i+1})$$

(22) $$\mathrm{var}(\eta_1) = \mathrm{var}(y_1)$$

(23) $$\mathrm{cov}(\eta_i, \eta_{i+1}) = \beta_{i+1,i}\mathrm{var}(\eta_i)$$

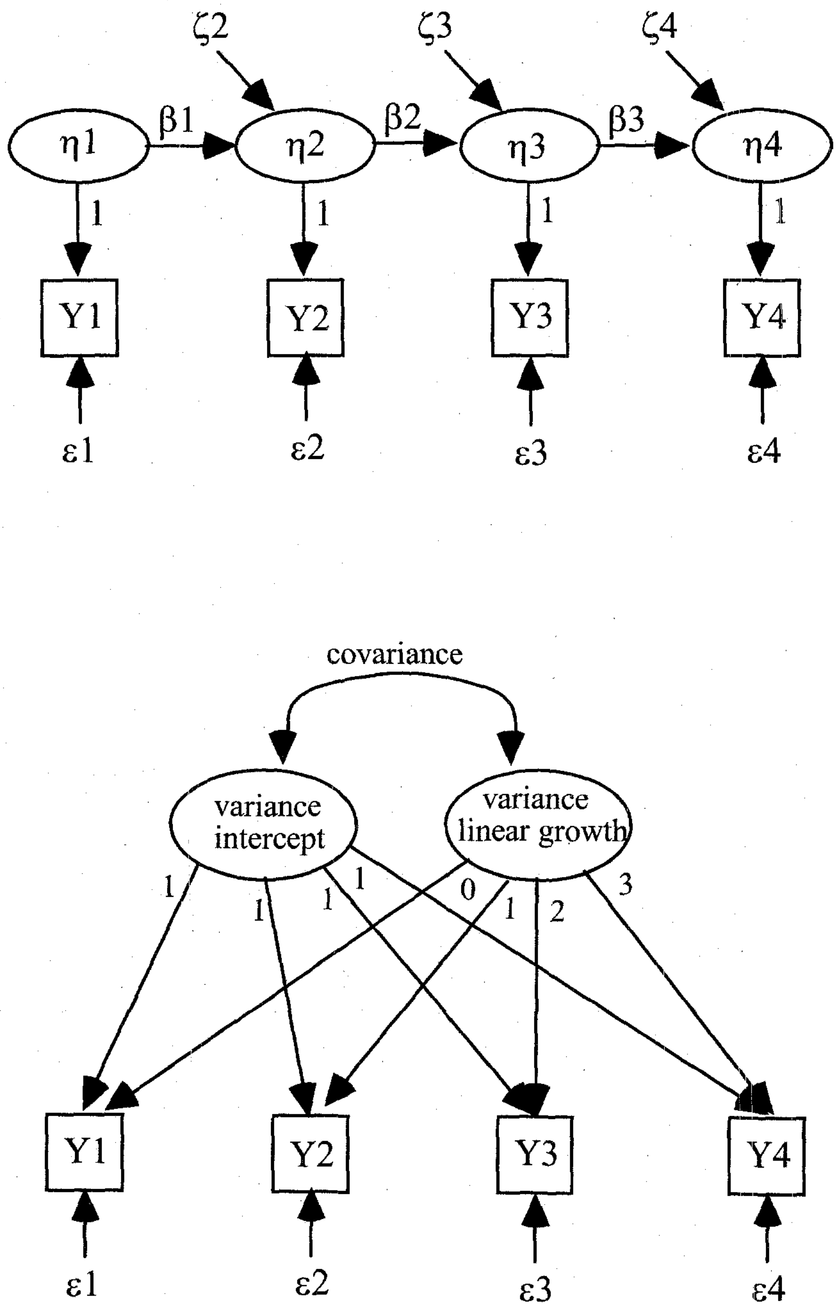


Figure 1
Examples of a Quasi Simplex Model (above) and a Linear Latent Growth Curve Model (below) for a Four Occasions Covariance Matrix (means excluded)

As can be derived from Equations 21 and 23 the stability of individual differences depends upon the relative magnitude of ζ_{i+1} in comparison with η_i, with perfect stability when ζ_{i+1} is zero. Furthermore, latent variance only increases when β^2_{i+1} is greater than one. In other words, the simplex model is not suited to represent the combination of increasing variance and perfect stability, but it can represent increasing variance in combination with high stability when the variance of ζ is relatively small.

Hypotheses concerning the stability of individual differences and increasing latent variance can be tested by comparing a model in which the $\zeta_{i+1,i}$ are constrained to be zero with a model in which these parameters are freely estimated. Hypotheses concerning changing influences can be tested by means of a multivariate model. In case of a multivariate simplex model lead lag effects between two different constructs, that is, the effects of latent variables on a previous time on other latent variables at a latter time, are specified. These lead lag effects are expected to explain a part of the variance of ζ, that is, the variance of the equation residuals. In the case of developmental limits, the lead lag effect between two constructs at time t will be significant, but will become insignificant at time t+1.

In contrast, in the LGC model growth is represented by a constant base of initial performance levels and changes of these initial levels as a function of time. Although the LGC model contains random changes in the sense of measurement error, it does not incorporate random disturbances as a part of the change in the latent scores (Kenny & Campbell, 1989). Therefore, the LGC model seems less suited to represent changing influences of the interindividual variance in reading. However, the model seems well suited to represent the expected finding of increasing interindividual variance in combination with perfect stability of the rank ordering of individuals. When novice readers with high initial reading performances also have the highest growth rates individual differences in reading will fan out with time. In terms of the LGC model this would mean that heterogeneity both in initial reading level and in rate of change across people, and a positive correlation between the initial level and the growth parameters is expected. The hypothesis of a positive correlation between initial reading level and growth rate can be tested by comparing a model in which the covariance between level and growth rate is estimated with a more restricted model in which one or all of the relevant elements of the Ψ matrix are constrained to be zero.

In case of the LGC model the requirement of changing influences can be represented in a cross-domain latent growth model (Stoolmiller, 1994; Willett & Sayer, 1996). In this model the latent growth curves describing the development of two different constructs are combined into one single model. In this manner, apart from the estimation of the latent growth parameters for both constructs, estimations of the correlation between these parameters can be found in matrix $\mathbf{\Psi}$. Correlations between initial status on one construct and growth rate on another construct now corresponds to the concept of lagged effects. When this correlation is positive, this would mean that the covariance between the constructs increases with time. When this correlation is negative the reverse is true. Hypotheses concerning the nature of the cross-domain relationships can be tested by comparing different models with various restrictions on the $\mathbf{\Psi}$ matrix.

To be able to test the assumption of common causation the mathematical model should analyze the means and covariance structure simultaneously. It is hypothesized that the factor loadings λ in case of the LGC model given by Equations 8 and 9, or the autoregressive coefficients β of the simplex model with structured means given by Equations 12 and 19 can account for time-dependent changes in both the longitudinal mean and the covariance structure (Mandys et al., 1994).

Because the LGC model seems to be the most appropriate model to represent the first aspect of Matthew effects in reading, and the Quasi simplex model must be preferred to represent the second aspect, empirical criteria are needed to decide upon the suitability of the two models. The analyses should therefore not only provide a quantitative description of the process and results for hypothesis tests but should also offer conceptual refinement and clarification on Matthew effects in reading.

Two Approaches to Data Analysis: Application from a Dutch Longitudinal Study in Reading

Methods

Sample

The present study tracked 235 children from 40 different schools through the early elementary grades. At the time of the first measurement the mean age of this sample was 74 months (*Sd* 4 months, *min* = 64 *max* = 88). Groups of students were excluded from the sampling procedure in order to arrive at a sample of students that could be tracked through the first three grades. Children at risk for non-promotion or referral to special education were excluded to minimize sample attrition. Children expected to be among the best readers of the group were excluded to prevent problems with the ability range of the measurement instruments. Once selected, students were tracked through subsequent classroom assignments effected by the schools' normal administrative procedures without influence from the research team.

The statistics of the selected group are based on the 235 subjects (121 boys and 114 girls) that completed the whole study. The selection of subjects was expected to yield a sample of students reading at or just above or under the average level for Grade 1, with less skilled readers being slightly over represented. It was also attempted to provide systematic variation in relevant prerequisite skills, age and gender in the final sample. Comparing the final sample with several reference groups reported in the test manuals, we conclude that the selection procedure has resulted in the intended sample of students.

Measures

In the present study reading ability is divided in two interconnected but distinct abilities specified as decoding ability and reading comprehension. A list of unrelated words (DMT, list 2; Verhoeven, 1992) was used as a repeated measure for decoding ability. The DMT2 was administered six times: three times in Grade 1, twice in Grade 2 and once at the end of Grade 3. Scoring consisted of the number of correct words read aloud within a time limit of one minute. For reading comprehension, two parallel versions of a test (BELL; Van den Bos, 1992) were used to assess the child's ability to understand the meaning of 39 unrelated sentences. This test was administered at the end of Grade 1, twice in Grade 2, and one time at the end of third Grade.

Statistical Analysis

Both the LGC and the Quasi simplex model with structured means were used to test three predictions of Matthew effects in reading: (a) fan-spread patterns when outcomes are plotted against time changing, (b) causal factors with reading development, and (c) the assumption of common causation. All analyses were performed within a structural equation modeling framework using the LISREL Version 8.0 computer program (Jöreskog & Sörbom, 1993). Throughout the present study estimation was carried out by mimimisation of the likelihood ratio function.

The LGC models and the Simplex models are compared on overall model fit. Two fit measures were used. The χ^2 statistic is a test of the null hypothesis of exact fit. Browne and Cudeck (1993) proposed to replace this hypothesis by a less implausible interval hypothesis of close fit. The hypothesis that the root mean square error of approximation (RMSEA) is less or equal to 0.05 is tested.

Within each type of model hypotheses are tested by making use of the sequential chi-square difference tests procedure. When the respecified model is nested under the first model, that is when its set of freely estimated parameters is a subset of those estimated in the first model, the null hypothesis of no significant difference between the two nested models can be tested. For instance, a linear latent growth model is nested within the quadratic model and the quadratic formulation within the cubic model. The difference between the χ^2 statistic values for nested models is itself distributed as χ^2, with degrees of freedom equal to the difference in degrees of freedom for the two models.

Results

Descriptives

Table 1 shows the sample mean vector and covariance matrix for the six occasion decoding measure DMT2 (**Y**1 - **Y**6), estimated with data on 235 children. The statistics in Table 1 display the expected increase of interindividual variability over time, judging by the leading diagonal of the covariance matrix, and a simplex structure of the data, as suggested by the decreasing covariances between occasions.

Table 2 shows the sample mean vector and covariance matrix for the four occasion reading comprehension measure BELL (**Y**3 - **Y**6), estimated

Table 1
Sample Mean Vector and Covariance Matrix for the Six Occasion of Decoding Ability

		Decoding Ability					
Statistic		**Y**1	**Y**2	**Y**3	**Y**4	**Y**5	**Y**6
Mean vector		1.368	13.657	25.037	49.610	55.992	70.981
Covariance matrix							
Y1	3 months	7.839					
Y2	7 months	5.612	66.129				
Y3	10 months	12.615	93.646	203.603			
Y4	17 months	12.289	93.308	201.393	340.221		
Y5	20 months	12.498	88.291	195.238	319.657	380.789	
Y6	30 months	8.729	68.624	156.543	276.030	313.554	331.622

with data on 235 children. The administration of reading comprehension measures started at the third measurement occasion at the end of Grade 1. The mean vector of the BELL shows a steady increase from the end of Grade 1 (Y3) to the end of Grade 3 (Y6). The statistics on the leading diagonal of the covariance matrix in Table 2 do not display systematically increasing interindividual variability over time. The structure of the covariance matrix tends to support a simplex structure of the data, as suggested by the decreasing covariances between occasions.

In Figures 2 and 3 the raw individual scores of decoding ability and comprehension in reading are plotted against time. Visual inspection of these plots suggests a linear growth model for both reading comprehension and decoding.

LGC versus Simplex Models: Fan-spread

First, several latent growth curve models were fitted to the covariance matrix described in Table 1. The time variable was expressed in months of instruction in reading received, with a summer time adjustment of two months, that is a school year consists of 10 months of instruction in reading. Measurements (Y1 - Y6) took place after respectively 3, 7, 10, 17, 20 and 30 months since the start of formal instruction in reading in Grade 1. To increase computational accuracy, the time variable was centered close to its mean (Prosser, Rasbash & Goldstein, 1991). This means that the first occasion in Grade 2 was chosen as the origin of time. In other words, the parameters λ_{i2} are set to respectively -14, -10, -7, 0, 3, and 13 ($i = 1, ..., 6$). Although the means and covariance structure can be analyzed

Table 2
Sample Mean Vector and Covariance Matrix for the Four Occasions of Reading Comprehension

	Reading Comprehension			
Statistic	Y3	Y4	Y5	Y6
Mean vector	11.888	17.093	17.758	23.383
Covariance matrix				
Y3 10 months	28.788			
Y4 17 months	11.150	35.337		
Y5 20 months	9.938	12.776	29.688	
Y6 30 months	6.275	9.283	10.881	29.092

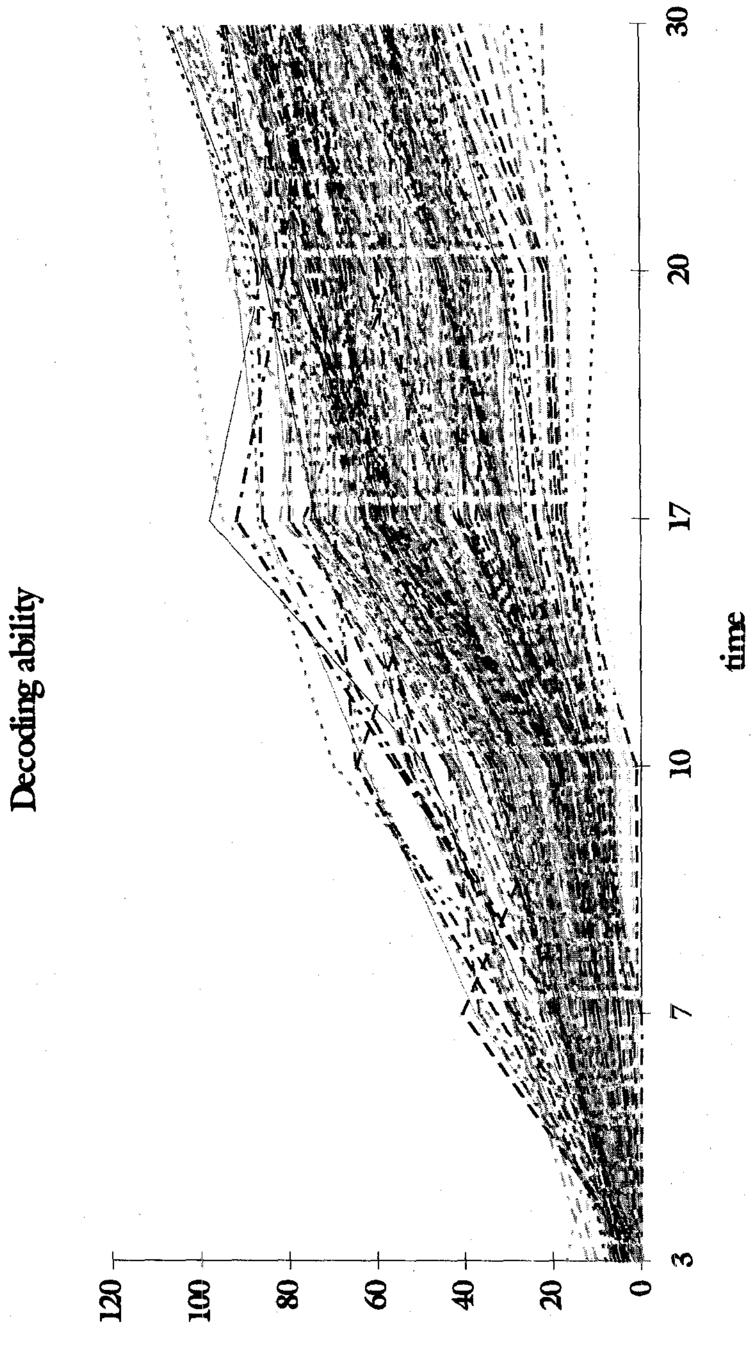


Figure 2
Raw Decoding Scores Plotted Against Number of Months of Instruction in Reading ($N = 235$)

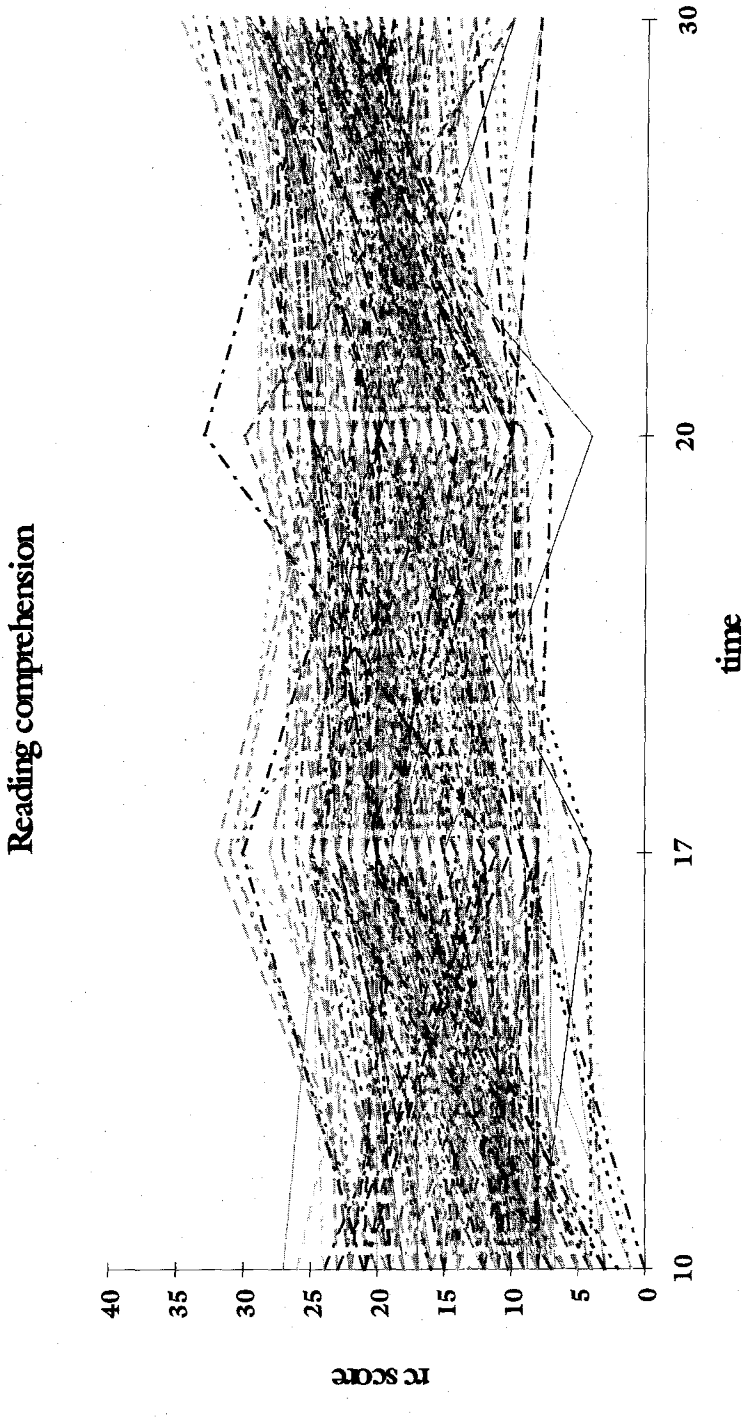


Figure 3
Raw Reading Comprehension Scores Plotted Against Number of Months of Instruction in Reading ($N = 235$)

simultaneously, the focus of the first hypothesis is on the structure and stability of individual differences. Therefore, a saturated model for the means was specified, that is the T means were modeled using T parameters in the first three models ($\boldsymbol{\alpha} = 0$, and $\mu_{yi} = \boldsymbol{\tau}_i$).

Table 3 shows the results of the latent growth curve models fitted to the decoding data. In this table the estimates of the parameters with their standard errors and two measures for overall model fit are reported. The parameters must be interpreted in the following manner: $\boldsymbol{\alpha}_1$ is the population average of the true decoding level after 17 months of instruction in reading, $\boldsymbol{\alpha}_2$ is the average linear growth after 17 months of instruction in

Table 3
Latent Growth Curve Models for Decoding Ability

Parameter	Model 1	Model 2	Model 3	Model 4
$\alpha 1$				48.117 (1.139)
$\alpha 2$				3.113 (.093)
$\alpha 3$				-.062 (.004)
$\alpha 4$				-.003 (.000)
$\psi 1$	277.344 (27.318)	290.831 (28.101)	282.357 (27.378)	290.206 (28.138)
$\psi 2$	.411 (.046)	1.663 (.195)	1.498 (.187)	1.637 (.196)
$\psi 3$	.002 (.000)	.003 (.000)	.003 (.000)	.003 (.000)
$\psi 4$		.000 (.000)	.000 (.000)	.000 (.000)
$\psi 1, 2$	8. 742 (.097)	12.858 (1.864)		12.777 (1.865)
$\psi 1, 3$	-.747 (.086)	-.815 (.089)	-.854 (.093)	-.810 (.089)
$\psi 1, 4$		-.022 (.007)	.038 (.004)	-.021 (.007)
$\psi 2, 3$	-.017 (.003)	-.029 (.006)	.012 (.003)	-.028 (.006)
$\psi 2, 4$		-.006 (.001)	-.007 (.001)	-.006 (.001)
$\psi 3, 4$		.000 (.000)	.000 (.000)	.000 (.000)
$\theta 1 = \theta 2$	18.196 (1.768)	10.297 (2.171)	13.796 (2.564)	10.896 (2.237)
$\theta 3 = \theta 4$	53.121 (4.148)	43.645 (3.649)	40.101 (3.527)	45.277 (3.769)
$\theta 5 = \theta 6$	50.461 (5.367)	24.675 (5.158)	29.682 (5.170)	27.304 (5.398)
χ^2 (*df*)	233.52 (12), $p < .001$	30.57 (8), $p < .001$	103.12 (9), $p < .001$	53.602 (10), $p < .001$
rmsea	.281, $p < .001$	.110, $p < .001$	.211, $p < .001$	.137, $p < .001$
difference		202.95 (4), $p < .001$	72.55 (1), $p < .001$	23.03 (2), $p < .001$

Note: The first three models were fitted to the covariance matrix only. Model 1 represents quadratic growth, model 2 is a polynomial to the third degree. In model 3 the covariance between the intercept and linear growth is fixed at zero. Model 4 addresses individual differences as well as the developmental mean trend.

reading, α_3 is the average quadratic component, and α_4 is the average cubic component; ψ_1 - ψ_4 represent the interindividual variance of these components; $\psi_{1,2}$ - $\psi_{3,4}$ represent the covariances between these components; and θ_1 - θ_6 represent error variance.

First, a quadratic growth model was fitted to the data (model 1). As can be seen in Table 3, the fit was poor as indicated by a significant model chi-square statistic and a significant value of RMSEA. Because the linear growth model is nested under the quadratic growth model, and the quadratic model under a third degree polynomial etcetera, the fit of these models can be compared with the sequential chi-square difference test procedure. The best fitting latent growth curve model turned out to be a polynomial of the third degree (model 2) with the following error structure: $\theta 1 = \theta 2$, $\theta 3 = \theta 4$, and $\theta 5 = \theta 6$. Models with homoscedastic ($\theta 1 = \theta 2 = \theta 3 = \theta 4 = \theta 5 = \theta 6$), heteroscedastic (all θs are freely estimated), or pairwise correlated heteroscedastic error structures (all θs are freely estimated and pairwise covariances between adjacent error terms are allowed) all fitted significantly worse or led to inadmissible solutions such as negative variance estimates.

Significant between student variation for all growth parameters and significant covariation between the growth parameters were found. Under this time specification, the variance of the intercept (ψ_1) and the variance of the linear component (ψ_2) represent the true variance in level and instantaneous growth rate after 17 months of instruction in reading. The variance of the quadratic (ψ_3) and cubic components (ψ_4) are characteristics of the entire growth trajectory. The correlation between the intercept and linear growth is

$$.58(\mathrm{cov}\psi_{1,2} \div \sqrt{\mathrm{var}\psi_1 \times \mathrm{var}\psi_2}),$$

indicating a moderate fan-spread pattern. Constraining the covariance between intercept and linear growth (model 3, $\mathrm{cov}\psi_{1,2} = 0$), that is, testing for the fan-spread pattern, lead to a significant deterioration in fit ($\Delta\chi^2 = 72.55$, $df = 1$, $p < .001$).

Next, the quasi simplex model with a saturated model for the means ($\boldsymbol{\alpha} = 0$, and $\mu_{y_i} = \boldsymbol{\tau}_i$) was used to test the same hypothesis for the decoding data. Table 4 shows the results for the simplex models. In these models $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ are parameters of the mean; $\beta_{i+1,i}$ represents the regression of $\boldsymbol{\eta}_{i+1}$ on $\boldsymbol{\eta}_i$; ψ represents the variance of the equation residuals; θ_1 - θ_6 represent error variance; $\mathrm{var}\boldsymbol{\eta}_1 = \psi_1$, and $\mathrm{var}\boldsymbol{\eta}_{i+1} = \beta^2_{i+1,i}\mathrm{var}\boldsymbol{\eta}_i + \psi_{i+1}$. The first model with the same error structure as with the LGC model, that is $\theta 1 = \theta 2$, $\theta 3 = \theta 4$, and $\theta 5 = \theta 6$ did not converge to a solution. Therefore, the model was respecified with zero error variances for the first and last occasion ($\theta 1 = 0$ and $\theta 6 = 0$).

Table 4
Simplex Models for Decoding Ability

Parameter	Model 1	Model 2
τ		-7.47 (.64)
α		8.85 (.62)
$\beta 2,1$	.74 (.15)	1.30 (.10)
$\beta 3,2$	2.17 (.13)	1.20 (.04)
$\beta 4,3$	.99 (.04)	1.44 (.04)
$\beta 5,4$	.96 (.04)	.96 (.01)
$\beta 6,5$	.86 (.03)	1.08 (.01)
$\psi 1$	7.84 (.72)	7.84 (.72)
$\psi 2$	38.83 (5.32)	61.56 (7.08)
$\psi 3$	0	0
$\psi 4$	134.30 (17.15)	125.96 (17.28)
$\psi 5$	57.54 (11.64)	38.43 (9.35)
$\psi 6$	61.40 (7.31)	63.84 (8.37)
$\theta 1$	0	0
$\theta 2$	22.99 (2.67)	16.23 (3.42)
$\theta 3$	.32 (7.65)	53.77 (6.65)
$\theta 4$	6.39 (9.16)	13.67 (7.93)
$\theta 5$	16.95 (6.40)	30.68 (5.82)
$\theta 6$	0	0
varη1	7.84	7.84
varη2	43.14	74.86
varη3	203.28	107.27
varη4	333.83	347.13
varη5	363.84	356.79
varη6	331.62	480.73
χ^2 (*df*)	7.92 (7), p=.340	160.67 (11), p<.001
rmsea	.024, p = .680	.240, p < .001
difference		152.75 (4), p < .001

Note: Model 1 was fitted to the covariance matrix only. Model 2 also adresses the mean trend.

The estimation of this model leads to a negative estimate for the variance of the equation residual of the third latent variable (ψ_3). This improper solution is probably caused by the fact that the value of the parameter in the population is very close to zero. In this situation a sample estimate may assume an inadmissible value due to sampling fluctuations (Bollen, 1989). Because the estimate for this variance did not depart significantly from zero this parameter was fixed at zero (model 1). The fit of this model is good. The probability of getting a chi-squared value larger than that actually obtained, given that the hypothesised model is true is $P = .340$.

Fan-spread patterns, that is the statistical equivalent of Matthew effects, consist of two components: increasing variance and consistency of interindividual differences. Only high stability of interindividual differences in combination with increasing variances indicates a fan-spread pattern, that is, increasing individual differences that preserve subjects' ordering. Non-crossing fan spread, that is perfect stability and increasing variance, is only obtained when all the random variances ψ are zero and all the βs are greater than one. It should be noted that in this case the simplex model reduces to a factor model. After constraining all other random variances ψ to zero model fit deteriorated significantly ($\Delta\chi^2 = 1237.87$, $df = 4$, $p < .001$). Thus, a model with zero random variances, except for ψ_3, had to be rejected. This means that a non-crossing fan-spread pattern could only be established from the second to the third measurement occasion ($\psi_3 = 0$ and $\beta_{3,2} = 2.17$). Although the latent variance increases for most of the other occasions (see varη_1 - varη_6 in Table 4), this increase is to a large extent due to the random variance ψ, resulting in changing rank orderings.

For reading comprehension the results of the LGC analyses are shown in Table 5. The administration of reading comprehension measures started at the third measurement occasion at the end of Grade 1, that is, after 10 months of instruction in reading. For comprehension in reading analysis started with a linear growth model. The parameters ψ_{i2} are set to -7, 0, 3, and 13 respectively ($i = 3, ..., 6$), with the following error structure: $\theta 1 = \theta 2$, and $\theta 3 = \theta 4$ (model 1). Under this specification $\boldsymbol{\alpha}_1$ is the population average of the true reading comprehension level after 17 months of instruction in reading, $\boldsymbol{\alpha}_2$ is the average linear growth after 17 months of instruction in reading; ψ_1 and ψ_2 represent the interindividual variance of these components; $\psi_{1,2}$ represents the covariances between these components; and θ_1 - θ_4 represent error variance.

Models with homoscedastic, heteroscedastic, or pairwise correlated heteroscedastic error structures did not result in significant improvements. Model fit could also not be improved by adding a quadratic growth component ($\Delta\chi^2 = 5.95$, $df = 3$, $p = .114$). The correlation between intercept

Table 5
Latent Growth Curve Models for Reading Comprehension

Parameter	Model 1	Model 2	Model 3	Model 4
$\alpha 1$				16.254 (.261)
$\alpha 2$				.561 (.021)
$\psi 1$	10.490 (1.524)	10.277 (1.435)	10.047 (1.432)	9.981 (1.433)
$\psi 2$	.020 (.012)	.018 (.012)	0	0
$\psi 1,2$	-.043 (.096)	0	0	0
$\theta 1 = \theta 2$	20.876 (1.725)	21.132 (1.730)	22.008 (1.691)	22.637 (1.734)
$\theta 3 = \theta 4$	17.739 (1.697)	17.687 (1.679)	19.346 (1.521)	19.347 (1.523)
χ^2 (*df*)	6.578 (5), $p = .254$	6.794 (6), $p = .340$	9.552 (7), $p = .215$	19.661 (9), $p = .020$
rmsea	.036, $p = .546$	.023, $p = .660$	.039, $p = .553$	.071, $p = .182$
difference		.216 (1), $p = .657$	2.758 (1), $p = .097$	10.109 (2), $p = .006$

Note: The first three models were fitted to the covariance matrix only. Model 1 represents linear growth. In model 2 the covariance between the intercept and linear growth is fixed at zero. In model 3 the variance of the linear growth parameter is set to zero. Model 4 addresses individual differences as well as the developmental mean trend.

and linear growth did not significantly differ from zero, as indicated by an insignificant chi-square difference with a model in which the covariance between intercept and linear growth is constrained to zero (model 2, cov$\psi_{1,2}$ = 0). Moreover, no significant variation in linear growth was found (model 3, $\psi_2 = 0$). In other words, no fan-spread pattern could be detected for the reading comprehension data.

Table 6 (next page) shows the simplex models fitted to the reading comprehension data. In these models $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ are parameters of the mean; $\beta_{i+1,i}$ represents the regression of η_{i+1} on η_i; ψ represents the variance of the equation residuals; θ_1 - θ_6 represent error variance; var$\eta_1 = \psi_1$, and var$\eta_{i+1} = \beta^2_{i+1,i}var\eta_i + \psi_{i+1}$.

The first model has the same error structure as the LCG model ($\theta 1 = \theta 2$, and $\theta 3 = \theta 4$). The estimation of this model leads to a negative estimate for the variance of the equation residual of the second latent variable ($\psi 2$). Because the estimate for this variance did not depart significantly from zero this parameter was fixed at zero (model 1). The fit of this model is excellent. However, as can be seen in Table 6 the standard errors of the random variances ψ_3 and ψ_4 are large compared to their estimates. Therefore, these variances were constrained to zero in the second model. The assumption of

Table 6
Simplex Models for Reading Comprehesion

Parameter	Model 1	Model 2	Model 3
τ			8.06 (.62)
α			3.76 (.42)
$\beta 2,1$	1.33 (.21)	1.34 (.23)	1.39 (.15)
$\beta 3,2$	.87 (.16)	.92 (.15)	.68 (.05)
$\beta 4,3$	.69 (.14)	.83 (.13)	1.17 (.07)
$\psi 1$	8.37 (2.33)	7.59 (2.20)	8.03 (2.12)
$\psi 2$	0	0	0
$\psi 3$	4.47 (1.36)	0	0
$\psi 4$	7.62 (1.68)	0	0
$\theta 1 = \theta 2$	20.44 (1.89)	21.41 (1.80)	20.85 (1.76)
$\theta 3 = \theta 4$	13.95 (3.29)	19.50 (1.65)	20.49 (1.62)
varη1	8.37	7.59	8.03
varη2	14.87	13.70	15.45
varη3	15.74	11.67	7.12
varη4	15.14	8.11	9.68
χ^2 (*df*)	0.23 (1), $p = .890$	4.95 (4), $p = .290$	10.63 (6), $p = .100$
rmsea	.003, $p = .940$	.032, $p = .560$	.057, $p = .350$
difference		4.72 (3), $p = .193$	5.68 (2), $p = .058$

Note: The first two models were fitted to the covariance matrix only. Model 3 also addresses the mean trend.

zero random variances proved to be tenable. Non-crossing fan-spread could only be established for the first to the second measurement occasion ($\psi_2 = 0$ and $\beta_{2,1} = 1.34$). After that individual differences are stable and the latent variance is decreasing ($\eta 3$ - $\eta 4$).

LGC versus Simplex Models: Developmentally Limited Causal Relationships

Differences in word recognition lead to differences in reading comprehension. Students who are not preoccupied with decoding issues can devote their full attention to the process of constructing meaning (Perfetti, 1985). However, word-recognition skills may account for relatively little of the variance once readers get beyond the beginning stages of reading. In other words, the causal relationship between decoding ability and comprehension in reading is hypothesized to be developmentally limited.

To test this hypothesis within a LGC modeling framework, the univariate growth models for decoding and reading comprehension without structured means were combined for the last four measurement occasions. To specify this model, the linear growth parameter λ_{i2} for decoding and the linear growth parameter λ_{i5} for reading comprehension are both set to -7, 0, 3, and 13 ($i = 1, ..., 4$). In contrast with the decoding models described in the previous paragraph, a linear growth model was estimated for decoding ability with $\theta 1 = \theta 2$, and $\theta 3 = \theta 4$. For reading comprehension all previously mentioned specifications were preserved, that is the variance of the linear growth parameter and every covariation of this parameter with other growth parameters is constrained to zero and $\theta 5 = \theta 6$, $\theta 7 = \theta 8$. Covariation between the growth parameters of decoding ability and comprehension in reading were allowed. The above specified model did not fit the data [$\chi^2 = 164.92(26), p < .001$]. Therefore, the quadratic growth component was added for the decoding model. After estimating the variance and covariances of the quadratic growth component model fit increased significantly ($\Delta\chi^2 = 111.69$, $df = 4, p < .001$).

Table 7 (next page) shows the results of the final multivariate LGC model. In this Table, the variance of the true decoding level after 10 months of reading instruction is represented by ψ_1, ψ_2 is the variance of the linear growth parameter for decoding after 10 months of instruction in reading; ψ_3 represent the variance of the quadratic growth parameter for decoding; ψ_4 represents the variance of the true reading comprehension level after 10 months of instruction in readings; ψ_5 represents the variance of the linear growth parameter for reading comprehension after 10 months of instruction in reading; the off diagonal elements represent the covariances between the growth parameters; θ_1 - θ_4 represent error variances for decoding; and θ_5 - θ_8 represent error variances for reading comprehension. Figure 4 summarises the final model in a path diagram.

The correlation between the intercepts of reading comprehension and decoding is

$$.64(\mathrm{cov}\psi_{1,4} \div \sqrt{\mathrm{var}\psi_1 \times \mathrm{var}\psi_4}),$$

indicating that good decoders are also good comprehenders. The correlation between the intercept of reading comprehension and linear growth of decoding is

Table 7
Multivariate Latent Growth Curve Models for Decoding Ability and Reading Comprehension

covariance matrix	$\psi 1$	$\psi 2$	$\psi 3$	$\psi 4$	$\psi 5$
$\psi 1$	301.38 (29.41)				
$\psi 2$	9.20 (1.40)	.77 (.11)			
$\psi 3$	-.83 (0.12)	-.04 (.00)	.03 (.00)		
$\psi 4$	36.26 (5.22)	.70 (.28)	-.09 (.02)	10.52 (1.52)	
$\psi 5$	0	0	0	0	0
$\theta 1 = \theta 2$	40.13 (4.79)				
$\theta 3 = \theta 4$	30.86 (4.71)				
$\theta 5 = \theta 6$	20.58 (1.57)				
$\theta 7 = \theta 8$	20.77 (1.58)				
x2 (*df*)		53.23 (22), $p < .001$			
rmsea		.077, $p = .042$			

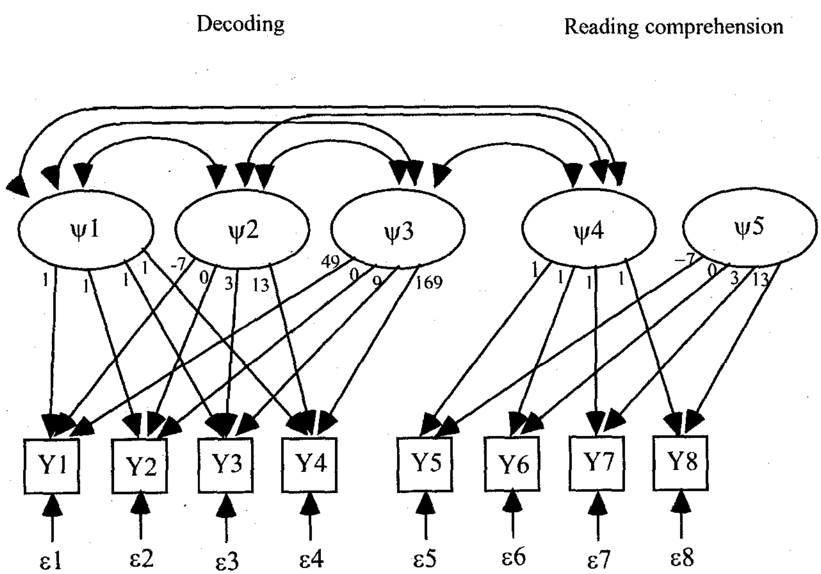


Figure 4
Path Diagram of a Multivariate Latent Growth Curve Model for Decoding and Reading Comprehension

$$.25(\text{cov}\psi_{2,4} \div \sqrt{\text{var}\psi_2 \times \text{var}\psi_4}),$$

indicating that children with faster growth rates for decoding ability will have higher future scores on reading comprehension.

Table 8 shows the results of the multivariate simplex models. The sequence of multivariate simplex models started with the specification of two separate simplex models for decoding ability and comprehension in

Table 8
Multivariate Simplex Models Decoding Ability and Reading Comprehension

Parameter	Model 1	Model 2
Decoding		
$\beta 2,1$	1.02 (.07)	1.04 (.07)
$\beta 3,2$	.96 (.04)	.97 (.04)
$\beta 4,3$	.86 (.03)	.86 (.03)
$\psi 1$	197.17 (20.39)	193.43 (20.79)
$\psi 2$	128.10 (22.09)	120.50 (21.38)
$\psi 3$	57.51 (11.64)	54.61 (11.40)
$\psi 4$	44.46 (11.57)	45.16 (11.60)
$\theta 1 = \theta 2$	6.42 (9.15)	10.14 (8.84)
$\theta 3 = \theta 4$	16.95 (6.40)	16.56 (6.40)
Reading Comprehension		
$\beta 6,5$	1.33 (.21)	1.33 (.25)
$\beta 7,6$	.87 (.16)	.89 (.19)
$\beta 8,7$	.69 (.14)	.67 (.16)
$\psi 5$	8.37 (2.33)	6.53 (2.21)
$\psi 6$	0	0
$\psi 7$	4.47 (3.29)	5.20 (3.64)
$\psi 8$	7.61 (4.52)	8.69 (4.99)
$\theta 5 = \theta 6$	20.44 (1.89)	21.65 (1.97)
$\theta 7 = \theta 8$	13.95 (3.29)	13.19 (3.61)
Lead-lag Relationships		
$\beta 6,1$		.08 (.02)
$\beta 7,2$		.00 (.02)
$\beta 8,3$		.00 (.02)
x2 (*df*)	105.09 (19), $p < .001$	95.37 (16), $p < .001$
rmsea	.139, $p < .001$	.146, $p < .001$
difference		9.72 (3), $p = .021$

reading. For reading comprehension the same model specification was used as described in the previous paragraph (model 1, Table 6), for decoding, in contrast with previous models, only the last four measurements were used. No cross-domain relationships were specified in the first model. Next, the lead-lag relations from decoding at time t to reading comprehension at time $t + 1$ were specified. Lead-lag relationships, that is the effects of latent variables on a previous time on other latent variables at a latter time, are considered to be suggestive of causal determination. The lead-lag relationships are represented by coefficient β (model 2). In other words, η_1 - η_4 represent the latent variances for decoding; η_5 - η_8 represent the latent variances for reading comprehension; $\beta_{i+1,i}$ represents the regression of η_{i+1} on η_i; ψ_1 - ψ_8 represents the variances of the equation residuals; and θ_1 - θ_8 represent error variance.

As can be seen in Table 8 only the effect of the first decoding variable on the second reading comprehension variable ($\beta_{6,1}$) was significant. The parameter estimates of all other effects were within the standard error of zero. Figure 5 summarises the final model in a path diagram. This result supports the notion of developmental limits in the relationship between decoding and comprehension in reading.

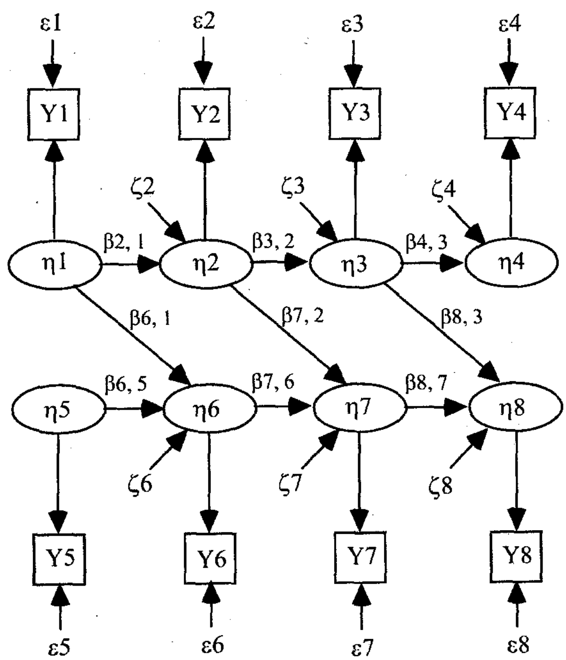


Figure 5
Path Diagram of a Multivariate Simplex Model for Decoding and Reading Comprehension

LGC versus Simplex Models: The Assumption of Common Causation

Finally, the hypothesis that the causes of variation between individuals are the same as the causes of variation in the means was tested. In case of the LGC model the factor loadings λ are assumed to account for time-dependent changes in both the longitudinal mean and the covariance structure. This hypothesis is tested by comparing a model with $\boldsymbol{\alpha} = 0$ and $\mu_{yi} = \boldsymbol{\tau}_i$ with a model in which $\boldsymbol{\tau}$ is constrained to zero and $\boldsymbol{\alpha}$ is estimated (see Equations 9 and 10).

Table 3 shows the results for decoding. The model with four $\boldsymbol{\alpha}$ parameters to model the mean trend (model 4) is compared with a model that uses six $\boldsymbol{\tau}$ parameters (model 2). As can be seen in Table 3 model fit deteriorated significantly ($\Delta\chi^2 = 23.03$, $df = 2$, $p < .001$) and therefore the assumption of common causation must be rejected. For comprehension in reading the results are shown in Table 5. Here, the model with two parameters to model the mean trend (model 4) is compared with a model with four $\boldsymbol{\tau}$ parameters (model 3). As can be seen in Table 5 the assumption of common causation had to be rejected ($\Delta\chi^2 = 10.11$, $df = 2$, $p < .006$).

In the Simplex model the assumption of common causation is tested by specifying a model in which the autoregressive coefficients β account for time-dependent changes in both the longitudinal mean and the covariance structure according to Equations 12 and 19. The results for decoding can be found in Table 4. In model 1 a saturated model for the means was used, that is $\mu_{yi} = \boldsymbol{\tau}_i$ and $\mu_{\eta i} = 0$, with six parameters to model the mean trend. The model with structured means, with two parameters to model the mean trend $\mu_{\eta 1} = \boldsymbol{\alpha}$, and $\mu_{\eta i+1} = \boldsymbol{\alpha} + \boldsymbol{\beta}_{i+1,i}\mu_{\eta 1}$ is nested under model 1. As can be seen in Table 4 the assumption of common causation is not tenable. However, as can be seen in Table 6 the assumption of common causation could not be rejected for reading comprehension. This last result indicates that both individual differences and the mean trend for comprehension in reading can be attributed to a common cause. A good candidate for such a cause is decoding ability.

Conclusions and Discussion

In this article three aspects of the Matthew effect hypothesis in reading have been tested by means of two distinct growth models. The Latent Growth Curve model and the Simplex model with structured means were used to represent fan-spread patterns, developmentally limited causal relationships, and the assumption of common causation in the development of decoding ability and comprehension in reading.

For decoding ability fan-spread is indicated by the results of the LGC model. The positive correlation of .58 between level and linear growth indicates divergence in decoding development. The results of the simplex models show that non-crossing fan-spread patterns could only be established for the second to the third measurement occasions, that is at the end of Grade 1. In Grade 2 and 3 latent variance increases with changing rank orderings. For comprehension in reading no fan-spread patterns were found for the LGC analyses. No individual variation for the linear growth parameters, and hence no covariation with level could be established. The results of the simplex models show fan-spread for the first to the second measurement occasions, that is at the beginning of Grade 2. After that time point the latent variance decreases.

Concerning developmentally limited causal relationships the results of a multivariate LGC model with decoding ability and comprehension in reading indicate that a relation between initial status and growth of the two constructs exist. Developmental limits in this relationship could not be established by means of the LGC, but were indicated by the results of the simplex models.

The assumption of common causation, that is, that the mean trend in reading skills and individual differences in reading abilities can be attributed to a yet unspecified common cause, had to be rejected for decoding ability. For comprehension in reading this assumption was tenable for the simplex model. Although in the case of the LGC the assumption of common causation had to be rejected on basis of the reported formal test, the deterioration in fit is not large and a visual inspection of the two models indicates that no large differences between the estimates for common parameters exist.

Whether the LGC model or the Simplex model must be preferred to tests different aspects of the Matthew effect hypothesis depends upon theoretical as well as empirical criteria. Both the correspondence between theoretical and formal representations of the model and the empirical fit to the data must be taken into consideration.

The LGC model and the Simplex model represent fundamentally different conceptions about the underlying growth process. The typical property of the simplex structure is that the sizes of correlations between measures collected at adjacent occasions are large and decrease systematically as a function of the number of occasions separating two repeated measures. This implies that the ordering of individuals or the size of differences between individuals will change as a function of time. Although this property of the simplex model corresponds to the concept of developmentally limited causal factors, the implication is that rank orderings are changing. In other words, perfect stability of individual differences can not be represented by the Simplex model. Contrary to the Simplex model the LGC model assumes no temporally related changes in the size of the correlations. The relation between initial status on the underlying construct and growth does not change with development. This conception of growth is in agreement with the fan-spread, but not with the assumption of changing causal factors. The issue of reciprocal causation has not been addressed in the presents study, but some suggestions can be made how to represent these bidirectional relations. In the LGC model bidirectional relations between variables can be represented by covariation between the growth parameters of the univariate growth models. However, because the reciprocal causal relations are hypothesised to be developmentally limited, the LGC model seems less suited. In the QSM, developmentally limited reciprocal causal relations can be represented by bidirectional lead-lag relationships between variables that are suggestive of causal determination. In sum, on basis of the formal representations of the three aspects of the Matthew model, no decision about which of the two models is more suited to represent all of these aspects can be made. Therefore, the empirical fit to the data must be taken into consideration.

Model fit seems to favor the Simplex model for decoding ability as well as comprehension in reading. The better fit of the simplex models is probably caused by the fact that the data seem to conform to a simplex structure. If the longitudinal data conform to a simplex structure the factor-analytic model is fundamentally unsuited (Wohlwill, 1973; Roskam, 1976; Boomsma & Molenaar, 1987). A more general reason for the lack of fit of the LGC model is that the model implies that the development is characterized by complete stability during the intermediate occasions. The relation between initial status on the underlying construct and growth does not change with development. Thus, only perfect stability of individual differences can be represented. However, this characteristic is in conflict with models of reading acquisition and probably also with most other developmental processes.

An additional empirical criterion, apart from overall model fit, is the comparison between the theoretical characteristics of the model and the observed characteristics of the data. With a first-order simplex the partial correlation between $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_{i+2}$ is zero when the intermediate occasion $\boldsymbol{\eta}_{i+1}$ is partialed out. The linear growth model implies that the correlation between $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_{i+2}$ is -1 when the intermediate occasion $\boldsymbol{\eta}_{i+1}$ is partialed out. For decoding ability the observed partial correlations were respectively corr($y1y3.y2$) = .20, corr($y2y4.y3$) = .01(n.s.), corr($y3y5.y4$) = .07(n.s.), and corr($y4y6.y5$) = .18. For comprehension in reading the observed partial correlations were corr($y1y3.y2$) = .23, and corr($y2y4.y3$) = .17. Although error variance affects these values, in general, these figures are more in agreement with the Simplex model.

In sum, the present research seems to favour the simplex model to represent Matthew effects in reading. However, single indicator quasi simplex models, as used in the present study, have come under attack in the literature for two reasons. First, the specification of correlated measurement errors is not possible within an one indicator simplex model. The assumption of uncorrelated measurement errors is likely to be false in the case of longitudinal data. As a result, relations among latent variables can be biased (see e.g. Marsh, 1993). With a multiple indicator quasi simplex model the assumption of uncorrelated measurement errors can be tested. Thus, the use of multiple indicators increases the power to detect specification errors in the model. However, the use of multiple indicators can introduce problems regarding factorial invariance. To formulate a growth model an important condition of the repeated observations has to be met. The individual skill that is supposed to change during development must retain a comparable meaning over the sequence of observations. Questions of whether the multiple measures used are measuring the same concept in all stages of development with the same unit of measurement and the same reliability must be tested empirically. When the test for factorial invariance is rejected, new problems can arise. Although we agree that in general a multiple indicator simplex model must be preferred above a single indicator model because in this manner the power of the model can be enhanced, in the present study no multiple indicators were used because no multiple measures were present in the data set. Moreover, the objective of the present study was to compare two models on the same data set.

Second, the simplex model is criticised for fitting to data sets that were generated by fundamentally different growth models, such as the linear growth model (Rogosa & Willett, 1985a). However, as shown by Mandys et al. (1994) the reverse is also true. Thus, empirically the two models are hard to distinguish (Rogosa & Willett, 1985a). This distinction becomes more clear when the number of measurement occasion increases (Mandys et al., 1994). In the present study, no large difference in fit between the LGC and the Simplex model could be found for the four occasion reading comprehension data. However, for the six occasion decoding data the difference in fit is significant.

An important objection with regard to the use of the LGC model can be made. The suitability of the LGC to uncover fan-spread rests on the assumption that the correlation between level and growth can be interpreted substantively. However, the value of this correlation is dependent on the origin of time. For instance, the correlation between level and linear growth of comprehension in reading is respectively -.37, -.09, and .04 when $t = 0$ for respectively the end of Grade 1, halfway Grade 2 and the end of Grade 2. Because there is no natural origin for time in most longitudinal studies interpretation of the correlation between status and growth is in most cases not valid (Rogosa & Willett, 1985b; Rovine & Molenaar, 1995). Moreover, also the interpretation of the variables related to the growth parameters is dependent upon the origin of time (see, for instance, Bryk & Raudenbush, 1989). For instance, the correlation between reading comprehension level and linear growth in decoding changes from .25 to .38 when $t3 = 0$ instead of $t4 = 0$.

Moreover, a useful practical aspect of the quasi simplex model is the fact that it reduces to a single factor model when the variances of ζ approach zero. If this happens to all these variances, the developmental model reduces to the single factor model presented by McArdle and Epstein (1987). If this happens at a number of intermediate adjacent measurement occasions (as is the case for the decoding data), the resulting model can be considered to be a QSM-LGC hybrid. From a theoretical point of view, this implies that the development is characterized by complete stability during the intermediate occasions. The total variances at these occasions may increase, decrease or remain constant depending on the value of the autoregressive coefficients. That the quasi simplex model can reduce to the latent growth curve model in this fashion testifies to the flexibility of this model.

On the basis of theoretical and empirical arguments mentioned above, we prefer the Simplex model to represent and analyze Matthew effects in reading. Because different factors are assumed to determine individual differences in reading ability at different developmental levels, a LGC model is not appropriate to describe growth in reading across developmental levels. A non-linear growth model such as the Simplex model in which, apart from the transmission of variance from occasion to occasion, new sources of interindividual variance can be incorporated seems to be more appropriate. Comparing the fit of these distinct models to the data supports this conclusion.

However, the results of the Simplex models imply that conceptual refinement and clarification on Matthew effects in reading are needed. The concepts of developmentally limited relationships and non-crossing fan-spread seem to be contradictory. A developmental pattern of increasing spread is only in agreement with changing causal factors when these factors change at the same time for all individual readers. For instance, a minor influence of word-recognition skills on comprehension in reading is expected after certain levels of word-recognition skills are reached.

However, it is not likely that individual readers reach this sufficient level of word recognition skills at the same time. As a consequence, the developmental curve of comprehension in reading for individual readers in a particular period of time can be caused by different factors, that is by the level of decoding skills in case of below average decoders and by language comprehension skills for above average decoders. Because of these different causal influences non-crossing fan spread patterns are not to be expected, but stability of individual differences can be high.

References

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Boomsma, D. I. & Molenaar, P. C. M. (1987). The genetic analysis of repeated measures I: Simplex models. *Behavior Genetics*, *17*, 111-123.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 136-162). London: Sage.

Bryk, A. S. & Raudenbush, S. W. (1989). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 159-204). New York: Academic Press.

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M. Kamil, P. Mosenthal, & P. D. Pearson, *Handbook of reading research. Volume II* (pp. 512-538). New York: Longman.

Hanna, G. & Lei, H. (1985). A longitudinal analysis using the Lisrel-model with structured means. *Journal of Educational Statistics, 10 (2)*, 161-169.

Hoover W. A. & Gough P. B. (1990). The simple view of reading. *Reading and Writing, 2*, 127- 160.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology*, *23*, 121-145.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303-351). New York: Academic Press.

Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7 — User's reference guide*. Chicago: Scientific Software International.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8 — User's reference guide*. Chicago: Scientific Software International.

Kenny, D. A. & Campbell, D. T. (1989). On the measurement of stability in over-time data. *Journal of Personality*, *57*, 445-481.

Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics*, *19 (3)*, 201-215.

Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex and one-factor models. *Journal of Educational Measurement*, *30 (2)*, 157-183.

McArdle, J. J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133.

McCall R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development*, *52*, 1-12.

Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

Merton, R. K. (1968). The Matthew effect in science. *Science*, *159*, 56-63.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Prosser, R., Rasbash, J., & Goldstein, H. (1991). *ML3 Software for three-level analysis: User's guide*. London: Institute of Education, University of London.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92 (3)*, 726-748.

Rogosa, D. R. & Willett, J. B. (1985a). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics*, *10 (2)*, 99-107.

Rogosa, D. R. & Willett, J. B. (1985b). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, *50 (2)*, 203-228.

Roskam, E. E. (1976). Multivariate analysis of change and growth: Critical review and perspectives. In D. N. M. de Gruijter & L. J. T. van der Kamp (Eds.), *Advances in psychological and educational measurement*. New York: Wiley.

Rovine, M. J. & Molenaar, P. C. M. (1995). *Inducing a correlation between level and shape in the latent growth curve model*. Manuscript submitted for publication.

Rudinger, G. & Wood, P. K. (1989). N's times, and number of variables in longitudinal research. In D. Magnusson & L. R. Bergman (Eds.), *Longitudinal research: Quality of data*. New York: Cambridge University Press.

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly*, *30*, 894-906.

Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360-407.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research*, *29 (3)*, 263-288.

Van den Bos, K. P. (1992). *De BELL-Test. Begrijpend lezen en -luisteren. Experimentele versie*. Groningen, The Netherlands: Rijksuniversiteit.

Verhoeven, L. (1992). *Drie-Minuten-Toets*. Arnhem, The Netherlands: Cito.

Walberg, H. J. & Tsai, S. (1983). Matthew effects in education. *American Educational Research Journal*, *20*, 359-373.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1978). A simplex model for analyzing academic growth. *Educational and Psychological Measurement*, *37*, 745-756.

Willett, J. B. & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116 (2)*, 363-381.

Willett, J. B. & Sayer, A. G. (1996). Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techiques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement*, *28 (1)*, 61-76.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.

Accepted February, 1997.